

Estimating the 4D respiratory lung motion by spatiotemporal registration and super-resolution image reconstruction

Guorong Wu^{a)}

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Qian Wang^{b)}

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599 and Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Jun Lian^{c)}

Department of Radiation Oncology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Dinggong Shen^{d)}

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

(Received 10 July 2012; revised 10 January 2013; accepted for publication 24 January 2013; published 25 February 2013)

Purpose: One of the main challenges in lung cancer radiation therapy is how to reduce the treatment margin but accommodate the geometric uncertainty of moving tumor. 4D-CT is able to provide the full range of motion information for the lung and tumor. However, accurate estimation of lung motion with respect to the respiratory phase is difficult due to various challenges in image registration, e.g., motion artifacts and large interslice thickness in 4D-CT. Meanwhile, the temporal coherence across respiration phases is usually not guaranteed in the conventional registration methods which consider each phase image in 4D-CT independently. To address these challenges, the authors present a unified approach to estimate the respiratory lung motion with two iterative steps.

Methods: *First*, the authors propose a novel spatiotemporal registration algorithm to align all phase images of 4D-CT (in low-resolution) to a high-resolution group-mean image in the common space. The temporal coherence of registration is maintained by a set of temporal fibers that delineate temporal correspondences across different respiratory phases. *Second*, a super-resolution technique is utilized to build the high-resolution group-mean image with more anatomical details than any individual phase image, thus largely alleviating the registration uncertainty especially in correspondence detection. In particular, the authors use the concept of sparse representation to keep the group-mean image as sharp as possible.

Results: The performance of our 4D motion estimation method has been extensively evaluated on both the simulated datasets and real lung 4D-CT datasets. In all experiments, our method achieves more accurate and consistent results in lung motion estimation than all other state-of-the-art approaches under comparison.

Conclusions: The authors have proposed a novel spatiotemporal registration method to estimate the lung motion in 4D-CT. Promising results have been obtained, which indicates the high applicability of our method in clinical lung cancer radiation therapy. © 2013 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4790689>]

Key words: 4D-CT, spatial-temporal registration, super-resolution, sparse representation, lung motion estimation

I. INTRODUCTION

Lung cancer is the most common cause of cancer-related death in men and women worldwide. Radiation therapy, using high-energy x rays to kill cancer cells, is widely used in lung cancer treatment. A general principle in radiation therapy is to confine the radiation field to the shape of tumor and to reduce the toxicity of radiation to the surrounding normal tissues. However, the respiratory motion incurs major uncer-

tainty in the planning and treatment delivery of radiation therapy of lung cancer.^{1,2}

I.A. Motivation

To ensure the dose coverage, a breathing motion margin is usually added to the clinical target volume (CTV) to define an internal target volume (ITV) on the planning CT. In the conventional 3D radiation therapy, these safety margins expand in

all directions empirically, which often significantly enlarges the ITV and might increase the risk of radiation injury on the surrounding normal structures. Recently, the advent of 4D-CT (with respiration phase as the fourth dimension) provides the full range of lung/tumor motion along respiratory phase. This helps the expansion of CTV in a more accurate and patient-specific way.^{3–5} However, 4D-CT does not provide respiratory motion information between consecutive phases explicitly, which boosts the investigation of deformable registration methods for estimating the respiratory motion from 4D-CT images. For example, one of the most important applications of 4D-CT deformable registration is to automatically determine the motion envelope of tumor.^{3,6,7} Thus, by using the estimated motion envelope to account for breathing motion, instead of simply expanding based on experience, the size of ITV can be greatly reduced. Other applications of 4D-CT deformable registration include ventilation imaging,⁸ motion-compensated cone-beam CT reconstruction,⁹ and respiratory motion modeling.¹⁰ With more application of 4D-CT, the 4D-CT deformable registration methods are expected to become more useful in the image-guided radiation therapy.^{1,5,7}

I.B. Related works

In the last decade, investigations have been conducted intensively to develop the methods for accurate estimation of the respiratory motion from 4D-CT by establishing temporal correspondences.^{11–19} For example, a reference volume can be aligned with each phase image in 4D-CT by the free-form registration method²⁰ to reveal the lung motion. Feature-based image registration methods for lung CTs can also be used.^{21,22} However, there are considerable limitations in these registration methods as detailed below:

- The first limitation comes from the independent registration of individual phase images toward the reference image. In this way, the temporal coherence in 4D-CT is completely discarded during the registration, and thus the temporal consistency of the estimated respiratory motions cannot be guaranteed. As a remedy, for example, Metz *et al.*¹³ extended 3D B-Spline to 3D + t B-Spline deformable model for registering all phase images of 4D-CT jointly.
- The second limitation comes from the potential bias of selecting a reference for aligning different phase images. For example, most methods use either the maximum inhale or the maximum exhale phase image as the reference to register all other phase images.^{12,18,23} To remove the potential bias caused by the arbitrary selection of the reference image, groupwise registration methods have been recently proposed in Refs. 13, 19, and 24 by registering all phase images to the common space simultaneously.
- The third limitation comes from the image quality of current 4D-CT. Since the modern CT scanner can only scan a limited region of human body at each couch position, the final 4D-CT has to be assembled by sorting multiple free-breathing CT segments with respect to the

couch position and tidal volume.^{25,26} However, because of the patient's free breathing during scan, CT segments at different couch positions cannot be scanned exactly at the same tidal volume. Thus, 4D-CT is often contaminated by motion artifacts which include motion blur, overlap and gapping.^{27–29} Although the motion-related artifacts could be reduced by the breath training,^{30,31} improved sorting method,^{1,29,32,33} utilization of internal anatomical features,^{25,34,35} and postimage processing on 4D-CT images,^{16,36,37} the interslice thickness (e.g., 3 mm), which is often much higher than the intraslice resolution (e.g., 1 mm), still remains as a big challenge in 4D-CT registration. Often, vessels become discontinuous across neighboring slices, which makes it difficult to perform reasonable registration especially for small vessels. Actually, this missing anatomical information in each phase image can be restored by integrating complementary information from the aligned phase images. For example, 4D-CT often acquires more than ten individual phase images, each with voxel resolution of $1.0 \times 1.0 \times 3.0$ mm. After registering them into the common space, for each 1 mm segment of lung (along superior–inferior direction), we have sufficient information (such as $\sim 10/3$ slices available) for reconstructing the high-resolution image for this particular lung segment, and thus we can eventually reduce the slice thickness in the group-mean image to 1 mm or even lower. The details of our proposed method are given below.

I.C. Our proposed method

To overcome these limitations, we present a novel registration framework to estimate lung respiratory motion in 4D-CT. Our method consists of two iterative steps. *In the first step*, a spatiotemporal registration algorithm is proposed to simultaneously align all phase images (in the original low resolution, particularly with large interslice thickness) onto a group-mean image (in high resolution) in the common space. In particular, we hierarchically select a set of key points to represent the shape of group-mean image, and then use these key points to guide the registration with the (low-resolution) phase images by robust feature matching. Note that the group-mean image will be reconstructed with more anatomical details than any phase image, by integrating the complementary information from the aligned phase images (as described in the second step below). Thus, the registration between the group-mean and every phase image can become relatively easy by taking advantage of the clear anatomy in the high-resolution group-mean image since each point in the low-resolution phase image can identify its correspondence in the group-mean image which might be difficult to achieve with respect to another low-resolution phase image. Meanwhile, by mapping the group-mean image onto the image domain of each phase, every key point in the group-mean shape has several warped points in different phase images, which can be assembled into a time sequence (with respect to respiratory phase) to form a virtual temporal fiber.³⁸ (It is worth noting that the temporal fiber we used here are totally different from the fibers in

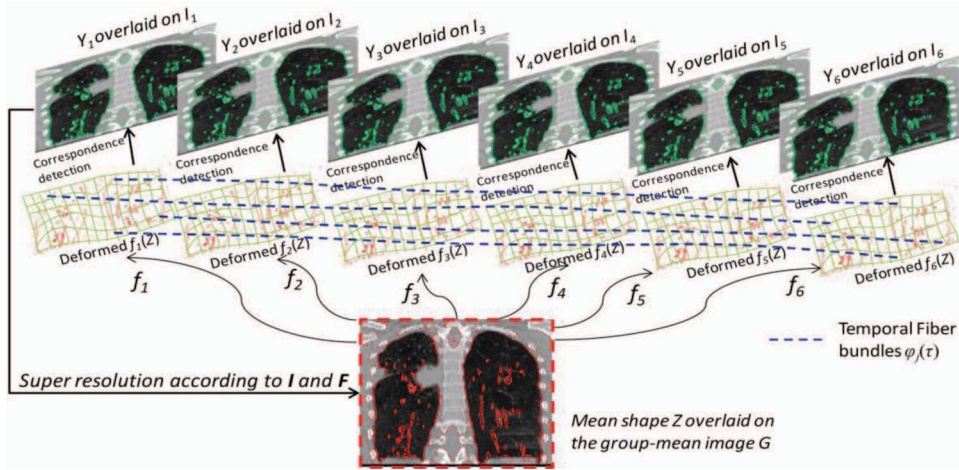


FIG. 1. Overview of our proposed method in respiratory motion estimation.

DT images.³⁹) Therefore, by requiring the smoothness on all these temporal fibers, the temporal coherence of 4D-CT registration can be assured.

In the second step, the high-resolution group-mean image is reconstructed by integration of the complementary information in all aligned phase images, by using a super-resolution technique developed in the computer vision area.^{40–42} Here, we assume that each local patch in the high-resolution group-mean image can be represented by the highly redundant local patches from those aligned phase images. Specifically, a l_0 -norm based sparse representation technique⁴¹ can be used to reconstruct the high-resolution group-mean image.

By alternatively repeating the above two steps, not only more accurate but also more consistent motion estimation along respiratory phase can be achieved for the 4D-CT than the widely used diffeomorphic Demons^{43,44} and the $nD + t$ B-Splines method,¹³ as confirmed by the experimental results on both simulated dataset and real lung dataset in Sec. III. In the following, we will present details of our motion estimation method for lung 4D-CT in Sec. II. The proposed method will be extensively evaluated and compared with the state-of-the-art 4D-CT registration methods in Sec. III. This paper will conclude in Sec. IV.

II. METHOD

We consider patient's lung as a 3D moving object during 4D-CT scan, and each phase image in 4D-CT is the respective 3D lung image taken at the particular respiratory phase. Thus, given the 4D-CT image $I = \{I_s | s = 1, \dots, N\}$ with N phases, we formulate the lung motion estimation as an unbiased groupwise registration problem under the constraint of temporal smoothness, to simultaneously register all phase images to the group-mean image G in the common space \mathcal{C} ,⁵⁹ i.e., $G \in \mathcal{C}$. It is worth noting that G is considered as a hidden variable in our method, which will be iteratively refined during the estimation of respiratory motion.

II.A. A Overview of our lung motion estimation method

In this paper, we will address two difficulties in respiratory motion estimation from 4D-CT: (1) how to preserve the temporal smoothness across respiratory phases without sacrificing the accuracy of motion estimation, and (2) how to overcome the dilemma of missing anatomy structures in each phase image because of the low image resolution in 4D-CT.

For the first problem, we use the similar strategy in our previous work³⁸ to maintain the spatiotemporal coherence in the framework of groupwise registration. Specifically, we simultaneously update the deformation field $f_s = \{f_s(x) | x \in \mathcal{C}\}$ for each phase image I_s with respect to the tentatively estimated group-mean image G in the common space \mathcal{C} . In general, G integrates all complementary information from the aligned I_s s and has higher resolution as detailed in Sec. II.C. In our method, we assume that each registered phase image I_s in the common space can be represented by a set of sparse points which are generated from a Gaussian mixture model (GMM).⁴⁵ The centers of GMM, called as the mean shape in this paper (as shown with points in the bottom of Fig. 1), are represented as $Z = \{z_j | z_j \in \mathbb{R}^3, j = 1, \dots, |Z|\}$, where z_j is the j th position of the mean shape in the common space. Then, for each z_j , its deformed positions in the image spaces of the respiratory phases $\{\tau_s, s = 1, \dots, N\}$ can be sequentially concatenated to form a continuous temporal fiber $\phi_j(\tau)$, which is a function of continuous respiration phase τ , with $\phi_j(\tau_s) = f_s(z_j)$ at discrete phase τ_s . In the middle of Fig. 1, we show the deformed group-mean shape in the space of each phase image I_s and the temporal fibers by the dashed curves. As we will make it clear in Sec. II.B, the heuristics of spatiotemporal coherence will be deployed along the temporal fiber bundles $\Phi = \{\phi_j(\tau) | j = 1, \dots, |Z|\}$.

In order to alleviate the ambiguities in image registration because of the missing anatomical structures, the super-resolution (SR) technique will be used to reconstruct a super-resolution group-mean image G based on all currently aligned phase images. In order to entitle the group-image G with both complete information and clear anatomical details, we use

the sparse representation technique⁴¹ to represent each local patch of the group-mean image by all candidate patches in the tentatively aligned phase images. In particular, l_0 norm regularization is used for selection of only a few candidate patches with non-zero weights for linear representation of each local patch in the group-mean image. This can be achieved by a greedy-fashion algorithm described below.

By iteratively repeating the spatiotemporal registration among all phase images and the super-resolution reconstruction on the group-mean image, all phase images will be registered to the common space, where the group-mean image G is reconstructed. It is worth noting that the main difference between our motion estimation method for lung 4D-CT and our previous work for longitudinal brain sequences³⁸ are: (1) The group-mean image G integrates complementary information from all phases by the super-resolution technique, thus it has much richer anatomical details than any phase image I_s alone. (2) The correspondence detection procedure is performed between the low-resolution phase image I_s and the super-resolution group-mean image G , thus each point in the low-resolution phase image I_s can easily find its correspondence in the super-resolution group-mean image with complete anatomical details, rather in any phase image with possible structure missing.

In the following, we will first describe the proposed spatiotemporal registration for 4D-CT in Sec. II.B. Then the super-resolution technique used in reconstruction of high-resolution group-mean image is presented in Sec. II.C. We summarize the whole method in Sec. II.D.

II.B. Hierarchical spatiotemporal registration of 4D-CT

In this section, we assume that the group-mean image G with super-resolution has been built in the previous round of registration (see Sec. II.C for detail).

Attribute vector: Before registration, we first segment each phase image into bone, soft tissue (muscle and fat), and lung.⁴⁶ Then, the vessels inside lung can be enhanced by multiscale Hessian filters⁴⁷ in order to allow the registration algorithm to focus on the alignment of lung vessels during registration. Thus, instead of using intensity only, we define an attribute vector $\vec{a}(y, I_s)$ as the morphologic signature for each point y in the phase image I_s to characterize the local image appearance. The attribute vector consists of image intensity, intensity gradient magnitude, and zero-order geometric moments⁴⁸ of each segmented structure (i.e., bone, soft tissue, lung, and vessel). Similarly, the attribute vector $\vec{a}(z, G)$ can be calculated for each point z in the group-mean image G . Note, each feature in the attribute vector is normalized from 0 to 1.

Key point selection: Given either the group-mean image or the individual phase image, two criteria are used for selecting key points: (1) key points should locate at distinctive regions since they will be relatively easy to identify during the correspondence detection; (2) key points should cover the entire lung image in order to derive the whole deformation in an accurate way, and the density of key points should be low in the uniform regions but high in the context-rich re-

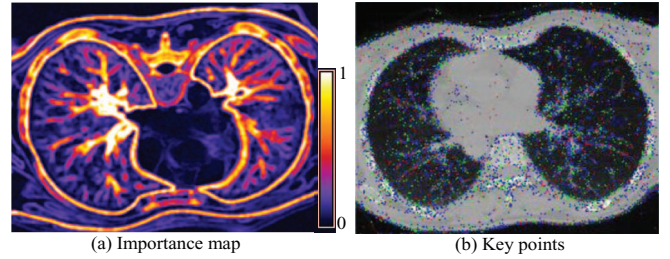


FIG. 2. The super-resolution algorithm used in our method.

gions. To meet these two criteria, we use the importance sampling strategy to hierarchically select key points. Specifically, we smooth and normalize the gradient magnitude values over the whole image domain. Then we use the obtained values as the importance (or probability) of each point to be selected as a key point during registration. Although advanced sampling method could be used here for guiding the key point selection, we use a simple but efficient strategy such as Monte Carlo simulation to sample key points, based on the importance map constructed. Figure 2 shows the nonuniform sampling based on the importance map [Fig. 2(a)]. The initial set of key points is displayed in red, and the key points added in the later stages of registration are displayed in green and blue in Fig. 2(b), respectively. It can be observed from Fig. 2 that the key points are more concentrated at context-rich (or edge-rich) regions, where the values of importance (or probability) are higher.

As the key points are usually located at the contexture edges, they can delineate the shapes of individual anatomical objects. Although the number of key points is much smaller than the image volume size, the selected key points are distributed in the entire image domain to steer the whole registration. Both key points in the group-mean image, $\mathbf{Z} = \{z_j | j = 1, \dots, |\mathbf{Z}|\}$, and in each phase image I_s , $\mathbf{Y}_s = \{y_i^s | i = 1, \dots, M_s\}$, are extracted. As we will explain next, we can decouple the complicated registration problem, with these key points, into two relatively simple subproblems: (1) robust correspondence detection on the key points; and (2) dense deformation interpolation from the correspondences established between key points.

II.B.1. Energy function for spatiotemporal 4D-CT registration

Considering that the respective shape in each phase image I_s can also be represented by its own key point set \mathbf{Y}_s , the key point y_i^s in the phase image I_s is regarded as an observation drawn from the GMM with its centers as the deformed mean shape \mathbf{Z} in the domain of the phase image I_s . Thus, given the deformation field f_s , the local discrepancy $\eta(z_j, y_i^s, f_s)$ between each z_j and y_i^s is defined as

$$\eta(z_j, y_i^s, f_s) = \beta \cdot \|f_s(z_j) - y_i^s\|^2 + (1 - \beta) \cdot \|\vec{a}(z_j, G) - \vec{a}(y_i^s, I_s)\|^2. \quad (1)$$

Obviously, it consists of both shape and appearance discrepancies, which are controlled by the parameter β . Specifically, in order to find the reliable anatomical correspondences, we

require: (1) the deformed mean-shape point z_j , i.e., $f_s(z_j)$, to be spatially close to the corresponding key point y_i^s in the phase image I_s ; (2) the attribute vector $\vec{a}(z_j, G)$ of the key point z_j in the group-mean image to be similar with the attribute vector $\vec{a}(y_i^s, I_s)$ of its corresponding point y_i^s in I_s . For robust correspondence matching, soft correspondence^{45,49,50} is allowed by using the spatial assignment $\pi_{i,j}^s$ to indicate the correspondence likelihood between key point y_i^s and particular z_j . It is worth noting that $\sum_{i=1}^{M_s} \pi_{i,j}^s = 1$.

Given $\mathbf{F} = \{f_s | s = 1, \dots, N\}$, we can describe the 4D-CT registration problem as a generative probability model, with $\{\mathbf{F}, \{\pi_{i,j}^s\}, \Phi\}$ as the model parameters and $\{y_i^s, a(y_i^s, I_s)\}$ as the observations. Then, the energy function for 4D-CT registration is inferred as

$$\begin{aligned} E(\mathbf{F}, \Phi, \{\pi_{i,j}^s\}) &= \sum_{s=1}^N \sum_{j=1}^{|Z|} \sum_{i=1}^{M_s} [\pi_{i,j}^s \cdot \eta(z_j, y_i^s, f_s) \\ &\quad + r \cdot \pi_{i,j}^s \cdot \ln(\pi_{i,j}^s)] + \sum_{s=1}^N L_s(f_s) \\ &\quad + \sum_{j=1}^{|Z|} L_T(\varphi_j(\tau)), \end{aligned} \quad (2)$$

where r is scalar, acting as the temperature in the annealing system. Apparently, the energy function consists of three terms. The first term describes the cost during the correspondence determination, which is performed between only a few key points in the phase images and in the mean shape of group-mean image. The second term measures the smoothness of each deformation field by $L_s(f_s)$, which is proportional to the bending energy of each f_s .^{45,51} The last term $L_T(\hat{\varphi}_j(\tau))$ measures the smoothness of each temporal fiber which is regulated by

$$L_T(\varphi_j(\tau)) = \sum_{s=1}^N [\varphi_j(\tau) - \varphi_j(\tau_s)]^2 \frac{1}{\sigma_\tau} \psi\left(\frac{(\tau - \tau_s) \equiv N}{\sigma_\tau}\right), \quad (3)$$

where ψ is the kernel regression function.⁵² Particularly, “ \equiv ” denotes the modulo operation since we consider lung motion forms a cyclical curve. In this paper we use the Gaussian kernel with bandwidth σ_τ . As we will see later, the advantages of using temporal fibers include: (1) The modeling of temporal motion is much easier on the particular temporal fiber $\varphi_j(\tau)$ than on the entire motion fields from f_s to f_{s+1} . (2) Spatial correspondence detection and temporal motion regularization are unified via temporal fibers.

II.B.2. Solutions of energy function for spatiotemporal 4D-CT registration

In general, it is very complicated to optimize \mathbf{F} , Φ , and $\{\pi_{i,j}^s\}$ in Eq. (2) jointly. Here we present an efficient solution in a divide-and-conquer way. Specifically, our solution to minimize energy function E consists of three steps. In the first step (\mathbf{SP}_1), since each spatial assignment $\pi_{i,j}^s$ is independent, we can estimate the degree of each $\pi_{i,j}^s$ by applying the

gradient descent based method to the terms in the bracket of Eq. (2). After removing those unrelated terms with \mathbf{F} and Φ , the remaining energy function turns to

$$\begin{aligned} E(\mathbf{F}, \Phi) &= \sum_{s=1}^N \sum_{j=1}^{|Z|} \sum_{i=1}^{M_s} \pi_{i,j}^s \|y_i^s - f_s(z_j)\|^2 + \sum_{s=1}^N L_s(f_s) \\ &\quad + \sum_{j=1}^{|Z|} L_T(\varphi_j(\tau)). \end{aligned} \quad (4)$$

However, the joint optimization of \mathbf{F} and Φ is still not trivial. Recall that $\varphi_j(\tau_s) = f_s(z_j)$ is the particular landmark on the temporal fiber $\varphi_j(\tau)$ at discrete phase τ_s . Then we can alternatively optimize \mathbf{F} and Φ as follows. We first minimize $E(\mathbf{F}, \Phi)$ with respect to $\varphi_j(\tau)$ by regarding $\varphi_j(\tau)$ as a whole and then find $\{\hat{f}_s\}$ which achieves the optimal $\hat{\varphi}_j(\tau_s)$ by solving the fitting problem. In this way, we have an updated relationship $\hat{f}_s(z_j) = \hat{\varphi}_j(\tau_s)$. Therefore, the individual objective functions for the two subproblems (\mathbf{SP}_2 and \mathbf{SP}_3) are given as

$$\mathbf{SP}_2: \min_{\varphi_j} \sum_{s=1}^N \sum_{j=1}^{|Z|} \sum_{i=1}^{M_s} \pi_{i,j}^s \|y_i^s - \varphi_j(\tau_s)\|^2 + \sum_{j=1}^{|Z|} L_T(\varphi_j(\tau)), \quad (5)$$

$$\mathbf{SP}_3: \min_{\mathbf{F}} \sum_{s=1}^N \left[\sum_{j=1}^{|Z|} \|\hat{\varphi}_j(\tau_s) - f_s(z_j)\|^2 + L_s(f_s) \right]. \quad (6)$$

\mathbf{SP}_2 aims to solve the spatiotemporal correspondence at time τ_s for a particular fiber $\varphi_j(\tau)$ as well as to estimate the continuous fiber $\hat{\varphi}_j(\tau)$ by kernel regression.⁵² \mathbf{SP}_3 is the problem of dense transformation field interpolation to obtain f_s for each phase image I_s . The solutions to \mathbf{SP}_1 – \mathbf{SP}_3 are provided below.

II.B.2.a. \mathbf{SP}_1 : Estimation of the spatial assignments. The spatial assignment $\pi_{i,j}^s$ can be calculated by minimizing E in Eq. (2) with respect to $\pi_{i,j}^s$ by requiring $\partial E / \partial \pi_{i,j}^s = 0$:

$$\pi_{i,j}^s = \exp\left\{-\frac{\eta(z_j, y_i^s, f_s)}{r}\right\} / \sum_{j=1}^{|Z|} \exp\left(-\frac{\eta(z_j, y_i^s, f_s)}{r}\right). \quad (7)$$

It is clear that $\pi_{i,j}^s$ is penalized in the exponential way according to the discrepancy degree $\eta(z_j, y_i^s, f_s)$ defined in Eq. (1). Notice that r is the denominator of the exponential function in Eq. (7). Therefore, when r is very high in the beginning of registration, even though the discrepancy between z_j and y_i^s might be large, the key point y_i^s still might have the chance to be selected as the correspondence of z_j with respect to I_s . As registration processes, the specificity of correspondence will be increased by gradually decreasing the temperature to a small degree, until only the single key point y_i^s with the smallest discrepancy being selected as the correspondences of z_j in the end of registration.

II.B.2.b. \mathbf{SP}_2 : Estimation of spatiotemporal correspondences. To this point, we are facing two coupled optimization problems in Eq. (5). One is the correspondence determination problem on z_j s, and the other is the estimation of continuous fiber $\varphi_j(\tau)$ along respiration phases which boils

TABLE I. Summary of parameters and their values used in the experiments.

β	The balance between shape and appearance discrepancy	$\beta = 0.5$
r	The temperature in soft correspondence assignment [Eq. (4)]	$r = 20 \times (1 - \delta)^2$
σ_τ	The kernel bandwidth in temporal smoothing [Eq. (8)]	$\sigma_\tau = 20 \times (1 - \delta)^2$
ε	The fitting error in super-resolution (algorithm shown in Fig. 4)	$\varepsilon = 5$
N_1	The search neighbor in constructing dictionary (used in Sec. II.C)	$5 \times 5 \times 5$
p	The size of local patch in each dimension (used in Sec. II.C)	$p = 3$

down to the kernel regression problem. The two optimization problems are difficult to solve altogether. However, the optimal solution to the correspondence at $\varphi_j(\tau_s)$ is much easier to compute if the entire fiber $\varphi_j(\tau)$ is fixed. As the result, we adopt an alternative optimization scheme. First, the spatial locations of the landmark points $\{\varphi_j(\tau_s) | s = 1, \dots, N\}$ sampled from fiber $\varphi_j(\tau)$ are obtained by optimizing the first term in Eq. (5) with respect to $\varphi_j(\tau_s)$ as

$$\varphi_j(\tau_s) = \sum_{i=1}^{M_s} \pi_{i,j}^s \cdot y_i^s. \quad (8)$$

Obvious, $\varphi_j(\tau_s)$ is the weighted mean location of all key points in the phase image I_s . After that, the continuous fiber $\hat{\varphi}_j(\tau)$ can be calculated by minimizing the energy $L_T(\varphi_j(\tau))$ defined in Eq. (3), which is widely known as the Nadaraya-Watson estimator:⁵³

$$\hat{\varphi}_j(\tau) = \frac{\sum_{s=1}^N \psi\left(\frac{(\tau - \tau_s) \equiv N}{\sigma_\tau}\right) \cdot \varphi_j(\tau_s)}{\sum_{s=1}^N \psi\left(\frac{(\tau - \tau_s) \equiv N}{\sigma_\tau}\right)} \quad (9)$$

Here, the kernel-based smoothing is performed along each fiber, to preserve the temporal smoothness. In general, high σ_τ results in smooth result at the expense of registration accuracy. In our implementation, we dynamically set the value of σ in the hierarchical way (see Table I): in the beginning, σ is relatively high to roughly determine the correspondences; with the progress of registration, we gradually decrease the value of σ_τ in order to achieve higher registration accuracy.

II.B.2.c. SP₃: Dense transformation interpolation. Since each deformation field f_s is independent to each other, we optimize f_s in Eq. (6) one by one in this step. For each f_s , it is required to minimize the fitting error of deformations on key points $\{z_j\}$ as well as the bending energy of entire deformation field, which is a TPS interpolation problem.^{45,51} Thus, by considering $\{z_j\}$ as the source point set and the deformations $\{\hat{\varphi}_j(\tau_s)\}$ in the particular phase image as the target point set, the dense deformation field f_s can be efficiently interpolated.

In Sec. II.C, we will introduce our super-resolution method for updating the group-mean image G with guidance of the estimated f_s s. After that, the group-mean shape Z can be extracted in the updated group-mean image G again to guide the next round of spatiotemporal registration.

II.C. Reconstruction of super-resolution group-mean image

Figure 3 shows the overview of our super-resolution algorithm, which is performed patch by patch. Here we use I'_s (displayed in the right of Fig. 3) to denote the aligned phase image I_s in the common space of the group-mean image, according to its transformation field f_s obtained in Sec. II.B. We first partition the whole space of the group-mean image G into a set of overlapping patches with patch size p in each dimension. Then, our goal is to obtain the super-resolution local patch $R_G(x)$ centered at $x \in \mathcal{C}$ (i.e., a box in the middle of Fig. 3) from an overcomplete dictionary of local patches, defined as

$$\mathbf{D} = \{D_t | D_t = R_{I'_s}(u), t = (u, s), s = 1, \dots, N, u \in \mathcal{N}_1(f_s(x))\}, \quad (10)$$

where $\mathcal{N}_1(f_s(x))$ denotes a small neighborhood at position $f_s(x)$, and $R_{I'_s}(u)$ is the local patch from the aligned phase image I'_s centered at location $u \in \mathcal{N}_1(f_s(x))$. It is clear that the dictionary \mathbf{D} consists of not only the local patch $R_{I'_s}(f_s(x))$ from each aligned phase image I'_s (designated by the pink boxes in the cross-sectional view and the pink lines in the superior/inferior direction), but also its surrounding patches within the neighborhood $\mathcal{N}_1(f_s(x))$ (designated by the blue boxes in the cross-sectional view and the blue lines in the superior/inferior direction). It is worth noting that we aim to enhance the resolution only along superior-inferior direction (which often has high interslice thickness), since the image quality of intraslice is already good enough with the voxel resolution such as 1×1 mm and also the motion in the short scanning time (< 0.1 s) for the particular cross-section slice can be ignored. Therefore, the local patch $R_G(x)$ to be enhanced and all entries in the dictionary \mathbf{D} are the cross-sectional 2D patches only.

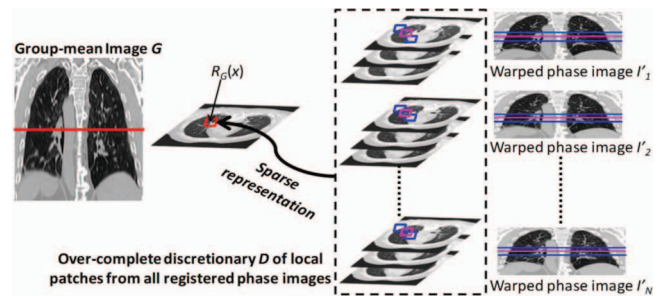


FIG. 3. The nonuniform sampling of key points using the importance (probability) map.

We regard that each $R_G(x)$ can be well represented by the local patches in the overcomplete dictionary \mathbf{D} . However, the use of all local patches for patch representation can only marginally improve the detail of the group-mean image, at the cost of adding more noise and blur in description of lung anatomy. Thus, in spirit of sparse representation, we use the l_0 regularization on the number of nonzero weights to guide the estimation of $R_G(x)$, which will be reconstructed by only a few candidate patches (with nonzero weights) in the dictionary \mathbf{D} . In this way, we keep the group-mean image as sharp as possible throughout registration. Specifically, we use $\mathbf{\Lambda}^K = \{D_{\chi(q)} | q = 1, \dots, K\}$ to denote the K elements with nonzero weights $\mathbf{w}(\mathbf{\Lambda}^K) = \{w_{\chi(q)} | q = 1, \dots, K\}$ selected from \mathbf{D} , where the indicator vector $\chi(q)$ identify the entry t [in Eq. (10)] of the q th selected local patch from the dictionary \mathbf{D} . Then the estimated patch $R_G(x) = \sum_{q=1}^K w_{\chi(q)} D_{\chi(q)}$ is eventually the sparse linear combination of all candidates in $\mathbf{\Lambda}^K$.

On the other hand, the estimated super-resolution local patch $R_G(x)$ should be close to the mean of all corresponding local patches of the aligned phase images, which can be defined as $\tilde{R}_G(x) = \frac{1}{N} \sum_{s=1}^N R_{I_s}(f_s(x))$. Thus, for estimating the super-resolution local patch $R_G(x)$, the energy function with respect to the weighting vector $\mathbf{w}(\mathbf{\Lambda}^K)$ can be defined as

$$\arg \min_{\mathbf{w}} \|\mathbf{w}(\mathbf{\Lambda}^K)_0\|, \quad \text{subject to } \theta(\mathbf{\Lambda}^K, \mathbf{w}(\mathbf{\Lambda}^K)) < \varepsilon,$$

$$\sum_{q=1}^K w_{\chi(q)} = 1,$$

where

$$\theta(\mathbf{\Lambda}^K, \mathbf{w}(\mathbf{\Lambda}^K)) = \left\| \tilde{R}_G(x) - \sum_{q=1}^K w_{\chi(q)} D_{\chi(q)} \right\|_2. \quad (11)$$

In general, the energy function in Eq. (11) uses the minimal number of local patches from dictionary \mathbf{D} to sparsely represent the local patch $R_G(x)$, with the fitting error (in l_0 -norm) less than ε . Here, we develop an efficient algorithm to drive the sparse linear combination $\sum_{q=1}^K w_{\chi(q)} D_{\chi(q)}$ in a greedy fashion. Similar to the Matching Pursuit algorithm,^{54,55} we sequentially select nonzero elements from dictionary \mathbf{D} . Suppose we have already selected k elements with nonzero weights, i.e., $\mathbf{\Lambda}^k = \{D_{\chi(1)}, D_{\chi(2)}, \dots, D_{\chi(k)}\}$ with the associated weighting vector $\mathbf{w}^k(\mathbf{\Lambda}^k) = \{w_{\chi(q)}^k | q = 1, \dots, k\}$, as defined above. It is worth noting that, for particular selected local patch $D_{\chi(q)}$, the value of $w_{\chi(q)}^k$ will be updated in each k th round, as indicated by the superscript k . Then, the current sparse representation of $R_G(x)$ is $\sum_{q=1}^k w_{\chi(q)}^k D_{\chi(q)}$. In the next step, we will examine every element from the remaining dictionary $\{\mathbf{D} - \mathbf{\Lambda}^k\}$ and continue to select $D_t \in \{\mathbf{D} - \mathbf{\Lambda}^k\}$ with the lowest degree of $\theta(\{\mathbf{\Lambda}^k + D_t\}, \mathbf{w}^{k+1}(\{\mathbf{\Lambda}^k + D_t\}))$ as $D_{\chi(k+1)}$, i.e.,

$$D_{\chi(k+1)} = \arg \min_{D_t \in \{\mathbf{D} - \mathbf{\Lambda}^k\}} \theta(\{\mathbf{\Lambda}^k + D_t\}, \mathbf{w}^{k+1}(\{\mathbf{\Lambda}^k + D_t\})). \quad (12)$$

Input: A local patch $R_G(x)$ and the over-complete dictionary \mathbf{D} .

Initialization: set $k = 0$, $\mathbf{\Lambda}^k = \{\}$ and $\varepsilon = 5$.

Loop: Iterate the following steps:

- For each entry $D_t \in \{\mathbf{D} - \mathbf{\Lambda}^k\}$, compute the weighting vector $\mathbf{w}^{k+1}(\{\mathbf{\Lambda}^k + D_t\})$ for the possible combination $\{\mathbf{\Lambda}^k + D_t\}$, by minimizing Eq. 13.
- Select the particular $D_t \in \{\mathbf{D} - \mathbf{\Lambda}^k\}$ with the minimal degree of $\theta(\{\mathbf{\Lambda}^k + D_t\}, \mathbf{w}^{k+1}(\{\mathbf{\Lambda}^k + D_t\}))$ as the new element $D_{\chi(k+1)}$.
- set $\mathbf{\Lambda}^{k+1} = \mathbf{\Lambda}^k \cup D_{\chi(k+1)}$, $\mathbf{w}^{k+1}(\mathbf{\Lambda}^{k+1}) = \mathbf{w}^{k+1}(\{\mathbf{\Lambda}^k + D_{\chi(k+1)}\})$ and $k = k + 1$.
- if $\theta(\mathbf{\Lambda}^k, \mathbf{w}^k(\mathbf{\Lambda}^k)) > \varepsilon$ and $\theta(\mathbf{\Lambda}^k, \mathbf{w}^k(\mathbf{\Lambda}^k)) < \theta(\mathbf{\Lambda}^{k-1}, \mathbf{w}^{k-1}(\mathbf{\Lambda}^{k-1}))$ then perform the next iteration. Otherwise, set $K = k$, $\mathbf{\Lambda}^K = \mathbf{\Lambda}^k$, $\mathbf{w}(\mathbf{\Lambda}^K) = \mathbf{w}^k(\mathbf{\Lambda}^k)$, and stop.

FIG. 4. The super-resolution algorithm in reconstructing one local patch by sparse representation.

For each candidate $D_t \in \{\mathbf{D} - \mathbf{\Lambda}^k\}$, the new weighting vector $\mathbf{w}^{k+1}(\{\mathbf{\Lambda}^k + D_t\})$ can be easily derived by solving the least-square problem below:

$$\begin{aligned} & \mathbf{w}^{k+1}(\{\mathbf{\Lambda}^k + D_t\}) \\ &= \arg \min_{\mathbf{w}^{k+1}} \left\| \tilde{R}_G(x) - \left(\sum_{q=1}^k w_{\chi(q)}^{k+1} D_{\chi(q)} + w_t^{k+1} D_t \right) \right\|_2^2, \\ & \text{subject to : } \sum_{q=1}^k w_{\chi(q)}^{k+1} + w_t^{k+1} = 1. \end{aligned} \quad (13)$$

The optimized weight $\mathbf{w}^{k+1}(\{\mathbf{\Lambda}^k + D_t\})$ in Eq. (13) can be computed in a closed form within three steps⁵⁶ (Appendix).

We can repeat this step of selecting an additional local patch from the remaining dictionary until the degree of $\theta(\cdot)$ in Eq. (11) is less than ε or starts to increase. The summary for estimating $R_G(x)$ from the overcomplete dictionary \mathbf{D} is listed in Fig. 4.

In our implementation, the whole group-mean image will be divided into a set of overlapping local patches with the size of each patch as 5×5 mm. After deploying the procedure of super-resolution reconstruction to all patches, we can obtain a super-resolution group-mean image G to better guide the spatiotemporal registration (Sec. II.B).

II.D. Summary

II.D.1. Implementation

The important parameters and the values in the experiments are listed in Table I. Note, some parameters automatically change their value with respect to the registration progress ratio δ ($0 \leq \delta \leq 1$). Besides the parameters listed in Table I, some other parameters in each image processing step are empirically determined as follows.

For each attribute vector, we calculate the zero-order geometric moments on each tissue type within a spherical neighborhood, where the neighborhood radius is set to 3 mm. Since we use nonuniform sampling strategy to select the key points, we need to set a threshold h to specify the percentile of key points in the size of brain volume (with background excluded), which is proportional with the registration progress ratio δ , i.e., $h = 0.01 + 0.09 \cdot \delta$. In order to save the

computation time, each key point z_j in the group-mean image only detects its correspondence in the particular phase image within a search neighborhood. In the beginning of registration, the search neighborhood is relatively large, i.e., with window size of $11 \times 11 \times 11$, and is linearly (with respect to δ) reduced to $3 \times 3 \times 3$ in the end of registration. After establishing the correspondence for each key point, TPS is deployed to produce the dense deformation field. Here, we follow the TPS interpolation method in Ref. 51. Specifically, since there are usually up to 10 000 of key points in a $256 \times 256 \times 100$ phase image, the complex matrix inversion in TPS is burdensome. We thus apply TPS interpolation block-by-block by using the overlapping blocks (i.e., with 1/3 overlap between neighboring blocks). Specifically, we set the block size to be $32 \times 32 \times 32$ and limit the number of key points in each block to be no more than 500 by uniform subsampling. It is worth noting that all of these parameters are fixed throughout all experiments.

II.D.2. Summary of the proposed method

Our registration-based lung motion estimation alternatively estimates the transformation fields F and refines the group-mean image G . The whole method is summarized as below:

1. Preprocess all phase images (e.g., histogram matching).
2. For each phase image, calculate the attribute vector for each point and extract the key points Y_s .
3. Use the phase image in the middle respiration phase as the initial group-mean image G by linearly interpolating it along superior–inferior direction, and then extract the mean shape Z .
4. Perform spatiotemporal registration for all phase images I , given the group-mean image G and the mean shape Z .
 - 4.1. Calculate the spatial assignment $\pi_{i,j}^s$ for each key point y_i^s with respect to each mean-shape point z_j in the current group-mean image [Eq. (7)].
 - 4.2. Deform each z_j to the phase image space I_s according to the spatial assignment obtained in step 4.1 [Eq. (8)].
 - 4.3. Perform the kernel smoothing along the temporal fibers [Eq. (9)].
 - 4.4. Interpolate the dense transformation field f_s by TPS.
5. Refine the group-mean image G by the super-resolution technique.
 - 5.1. Warp each phase image to the common space by F .
 - 5.2. Divide the whole image into a set of overlapping local patches and then obtain the high-resolution local patch $R_G(x)$ according to the super-resolution algorithm summarized in Fig. 4.
 - 5.3. Extract the mean shape Z from the latest-updated group-mean image G .
6. If the stop criterion (e.g., beyond the specified iteration number) is not satisfied, go to Step 4. Otherwise, output the estimated transformation fields F and

the group-mean image G . After completing the registration, for each point $x \in G$, its respiratory motion can be described by the respiratory motion function $\varphi(x, \tau)$, with $\varphi(x, \tau) = f_s(x)$ at discrete phase τ_s .

III. EXPERIMENTS AND DISCUSSION

To demonstrate the performance of our proposed registration algorithm in estimating lung motions, we evaluate its accuracy on both simulated data and real DIR-lab patient data,¹¹ by comparison with the popular pairwise diffeomorphic Demons⁴³ and the $nD + t$ B-Splines registration algorithm.¹³ (In the following experiments, the iteration numbers are set to 100, 100, and 100 for low, middle, and high resolution, respectively. The sigma value for smoothing deformation field is set to 2.2. For other parameters, we use the default settings.) For fair comparison, we use diffeomorphic Demons with ITK (www.itk.org) (Insight Segmentation and Registration Toolkit) based source code downloaded from <http://www.insight-journal.org/browse/publication/154>. To perform B-Spline based $nD + t$ registration algorithm on 4D-CT, we first compile the source code of “elastix” toolbox⁵⁷ (<http://elastix.isi.uu.nl/index.php>) and then use the parameter file provided by authors at <http://elastix.bigr.nl/wiki/index.php/Par0012>.

III.A. Evaluation on simulated dataset

III.A.1. Simulation of 4D-CT dataset

To validate the performance of our registration based motion estimation method, we simulate 4D-CT dataset as follows. The simulation starts from several 4D-CT images selected from DIR-lab dataset,¹¹ each of which contains six phases (τ_1, \dots, τ_6) with voxel size $0.97 \times 0.97 \times 2.5$ mm. For each 4D-CT image, we set the maximum inhale phase image I_1 at phase τ_1 as the reference image and use B-Spline based registration algorithm^{57,58} to align all other phase images onto the reference image. Then, we randomly add noises to B-Spline control points with ± 2 mm perturbation, in order to simulate the ground-truth deformation fields. The next step is to generate the simulated 4D-CT. Specifically, we first invert these ground-truth transformation fields and then use them to deform I_1 to the space of all other phases for obtaining a new simulated 4D-CT, i.e., with new phase images $\tilde{I} = \tilde{I}_1, \dots, \tilde{I}_6$. In this way, we can get the simulated 4D-CT \tilde{I} (with voxel size $0.97 \times 0.97 \times 2.5$ mm), along with the ground-truth deformation fields $\tilde{F} = \{\tilde{f}_{1 \rightarrow 1}, \dots, \tilde{f}_{1 \rightarrow 6}\}$ from \tilde{I}_1 to all other phase images of \tilde{I} . It is worth noting that $\tilde{I}_1 = I_1$ and $\tilde{f}_{1 \rightarrow 1}$ is the empty deformation field with displacement zero everywhere. Finally, we use Gaussian sampling to obtain the down-sampled 4D-CT in two steps: (1) first convolve each phase CT only along the superior–inferior direction with a one-dimensional Gaussian kernel $\frac{1}{16}[1 \ 4 \ 6 \ 4 \ 1]$; (2) then remove every even-numbered slice. After that, we use $\tilde{I}_\downarrow = \{\tilde{I}_{s\downarrow} | s = 1, \dots, 6\}$ to represent the down-sampled 4D-CT with lower voxel resolution of $0.97 \times 0.97 \times 5.0$ mm. Since B-Spline based registration algorithm is already used for generating simulated 4D-CT, we compare the performance

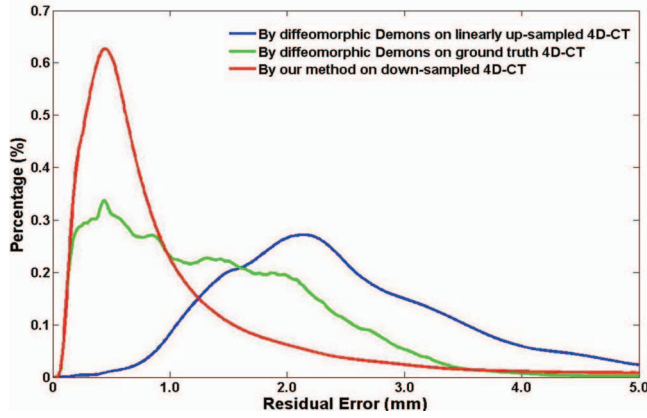


FIG. 5. The distributions of residual errors by different registration methods.

of our motion estimation method on simulated dataset only with diffeomorphic Demons, in order to avoid the bias of deformation model.

III.A.2. Experiment setup

In this experiment, we compare two registration methods (diffeomorphic Demons and our proposed method) on the simulated 4D-CT datasets. For diffeomorphic Demons, we linearly interpolate the down-sampled 4D-CT by increasing voxel resolution from $0.97 \times 0.97 \times 5.0$ mm to $0.97 \times 0.97 \times 2.5$ mm before registration, since the ground-truth 4D-CT has voxel resolution $0.97 \times 0.97 \times 2.5$ mm. Then, we use diffeomorphic Demons to register the linear-interpolated $\tilde{I}_{1\downarrow}$ with all other linear-interpolated phase images, obtaining $\mathbf{F}^{\text{Demons_L}} = \{f_{1 \rightarrow s}^{\text{Demons_L}} | s = 1, \dots, 6\}$. Furthermore, we apply diffeomorphic Demons to register ground truth data \tilde{I}_1 with all other ground truth data $\tilde{I}_2 \sim \tilde{I}_6$, obtaining $\mathbf{F}^{\text{Demons_G}} = \{f_{1 \rightarrow s}^{\text{Demons_G}} | s = 1, \dots, 6\}$. On the other hand, we apply our method directly to the down-sampled 4D-CT \tilde{I}_{\downarrow} (with low voxel resolution $0.97 \times 0.97 \times 5.0$ mm) and obtain the transformation fields $\mathbf{F} = \{f_1, \dots, f_6\}$ (directing from the common group-mean image space to each phase image), as well as the super-resolution group-mean image G (with voxel resolution $0.97 \times 0.97 \times 2.5$ mm). Since the ground-truth transformation field is defined in the space of \tilde{I}_1 , we redefine our obtained transformation fields by connecting the respec-

tive deformation field with the inverse transformation field of f_1 , i.e., $\mathbf{F}^{\text{Ours}} = \{f_{1 \rightarrow s}^{\text{Ours}} | f_{1 \rightarrow s}^{\text{Ours}} = f_1^{-1 \circ} f_s, s = 1, \dots, 6\}$.

In this way, we can evaluate the performance of our method in (1) estimating the respiratory motions, by comparing with diffeomorphic Demons, and (2) reconstructing the high-resolution image from the down-sampled phase images, by comparing with the simulated ground-truth 4D-CT.

III.A.3. Accuracy of respiratory motion estimation

We compare the residual errors between the ground-truth transformation fields $\tilde{\mathbf{F}}$ and the estimated transformation fields by the three different registration ways ($\mathbf{F}^{\text{Demons_L}}$ by diffeomorphic Demons on linearly interpolated 4D-CT, $\mathbf{F}^{\text{Demons_G}}$ by diffeomorphic Demons on ground truth 4D-CT, and \mathbf{F}^{Ours} by our method on low-resolution 4D-CT \tilde{I}_{\downarrow}). The average and standard deviation (STD) of residual errors are 2.03 ± 1.58 mm with respect to $\mathbf{F}^{\text{Demons_L}}$, 1.26 ± 1.15 mm with respect to $\mathbf{F}^{\text{Demons_G}}$, and 1.07 ± 1.47 mm with respect to \mathbf{F}^{Ours} , with all error distributions also shown in Fig. 5. Furthermore, we display the residual error on the lung surface by these three different registration ways through Figs. 6(a)–6(c). Apparently, our method achieves even slightly better registration accuracy than that the diffeomorphic Demons on the *ground-truth 4D-CT*.

III.A.4. Quality of estimated high-resolution lung images

By deforming the super-resolution group-mean image G built in our method to each phase image space, we can obtain the resolution-enhanced phase images. In Fig. 7, we can visually examine one typical case (at maximum exhale phase) of the simulated 4D-CT \tilde{I} [Fig. 7(a)], the linear-interpolated images from the down-sampled 4D-CT \tilde{I}_{\downarrow} [Fig. 7(b)], and the resolution-enhanced images by our method [Fig. 7(c)], with the intensity difference maps indicating the substantial improvement in anatomical details by our method.

III.B. Evaluation on DIR-lab dataset

There are ten cases in DIR-lab dataset, each of which has a 4D-CT with six phases. The intraslice resolution is around 1×1 mm, and the slice thickness is 2.5 mm. For each case,

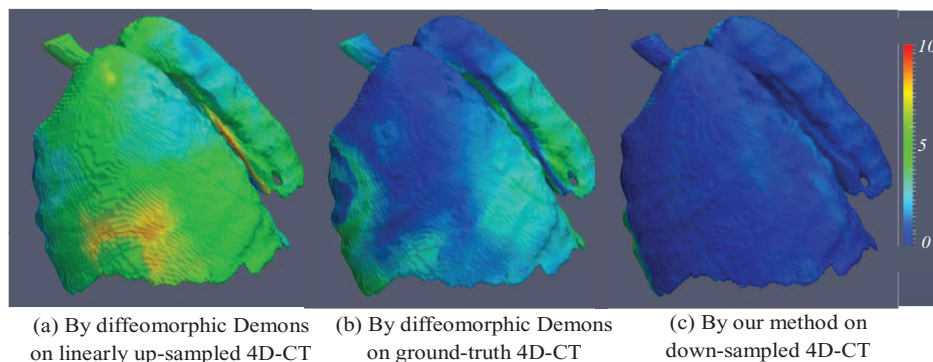
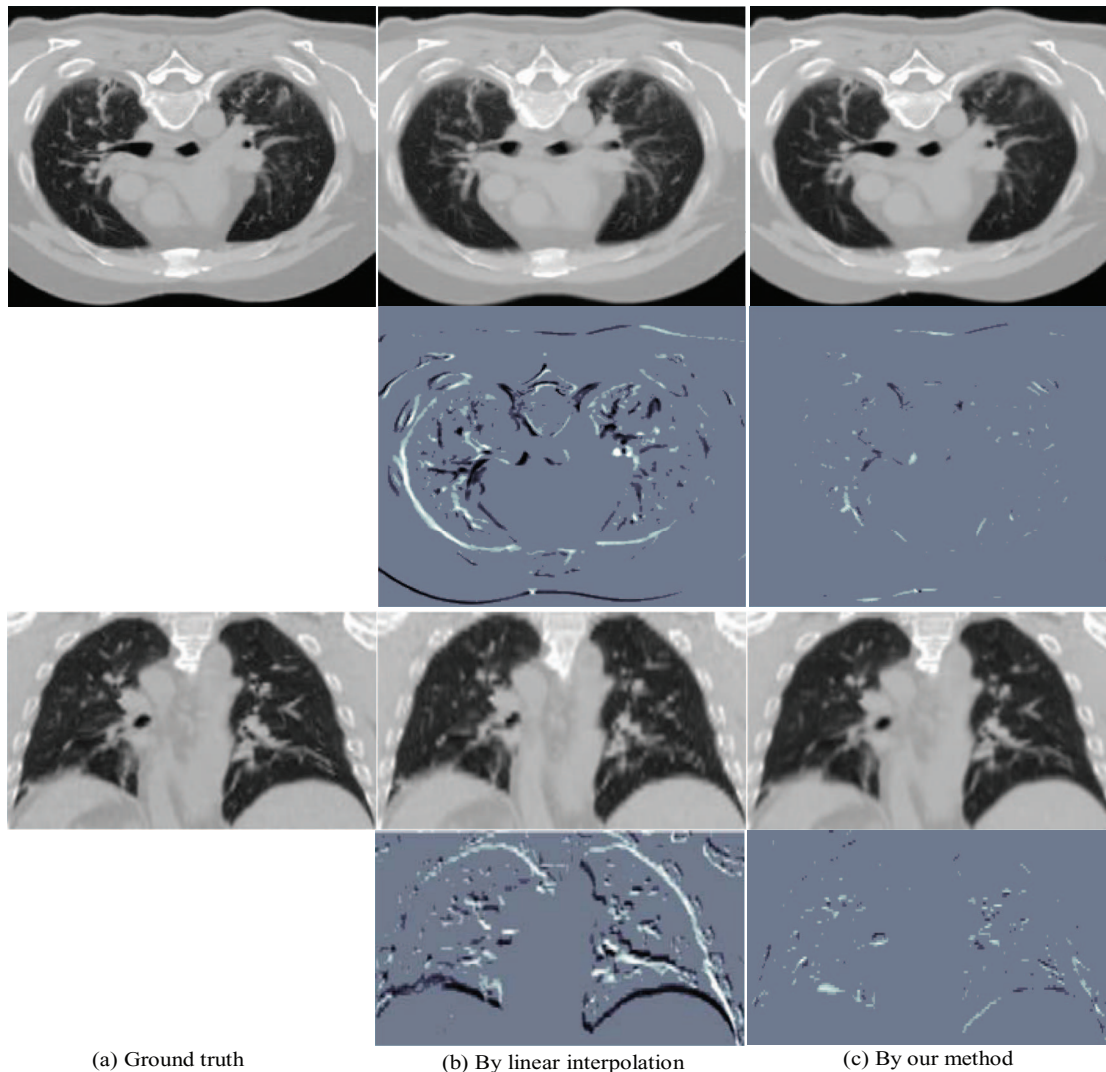


FIG. 6. The residual errors with respect to ground-truth on the lung surface by different registration methods.



(a) Ground truth (b) By linear interpolation (c) By our method

FIG. 7. Resolution-enhanced images at the maximum exhale phase by our method and linear interpolation.

300 corresponding landmarks in the maximum inhale (MI) and the maximum exhale (ME) phases are manually delineated. Also, correspondences of 75 landmarks are provided for each phase. Thus, the registration accuracy can be evaluated by the Euclidean distance between the reference landmark positions and the landmark positions propagated from the first phase to all other phases by registration algorithm. It is worth noting that our registration is performed on the entire 4D-CT, regardless of the evaluation on the 300 landmarks in MI and ME phases or on the 75 landmarks in all six phases. The voxel resolution in our group-mean image is enhanced to $1.0 \times 1.0 \times 1.0$ mm.

III.B.1. Accuracy of respiratory motion estimation on DIR-lab dataset

The registration results by diffeomorphic Demons, $nD + t$ B-Splines, and our method on the 300 landmarks between MI and ME phases are shown in Table II. Here, the MI phase image is selected as the reference image for diffeomorphic Demons. Note that the results by $nD + t$ B-Splines method are only available for the first five cases.¹³ (Although we

have the source code and parameter files for B-Spline based 4D registration method, this method needs the morphologically dilated lung mask during registration. Thus, we only cite registration error of cases 1–5, which the authors provide in

TABLE II. The mean and standard deviation of registration errors (mm) on 300 landmark points between maximum inhale and exhale phases. Bold is used for emphasis that the residual errors by our method are significantly smaller than other counterparts.

#	Initial	300 landmark points between MI and ME		
		Diffeomorphic Demons	$nD + t$ B-Splines	Our method
1	3.89 ± 2.78	1.82 ± 1.42	1.02 ± 0.50	0.64 ± 0.61
2	4.34 ± 3.90	1.87 ± 1.30	1.06 ± 0.56	0.56 ± 0.63
3	6.94 ± 4.05	2.30 ± 1.06	1.19 ± 0.66	0.70 ± 0.68
4	9.83 ± 4.85	4.31 ± 2.47	1.57 ± 1.20	0.91 ± 0.79
5	7.48 ± 5.50	1.96 ± 1.44	1.73 ± 1.49	1.10 ± 1.14
6	10.89 ± 6.97	6.23 ± 4.06	...	3.28 ± 3.45
7	11.03 ± 7.42	4.79 ± 3.10	...	1.68 ± 1.22
8	14.99 ± 9.01	6.17 ± 4.14	...	1.70 ± 1.69
9	7.92 ± 3.98	3.91 ± 1.97	...	1.72 ± 1.32
10	7.30 ± 6.35	3.47 ± 2.85	...	1.48 ± 1.84

TABLE III. The mean and standard deviation of registration errors (mm) on 75 landmark points across all six phases. Bold is used for emphasis that the residual errors by our method are significantly smaller than other counterparts.

#	Initial	75 landmark points across all six phases		
		Diffeomorphic Demons	$nD + t$ B-Splines	Our method
1	2.18 ± 2.54	1.10 ± 0.80	0.95 ± 0.65	0.51 ± 0.39
2	3.78 ± 3.69	1.42 ± 0.93	1.00 ± 0.62	0.47 ± 0.34
3	5.05 ± 3.81	1.34 ± 0.63	1.14 ± 0.61	0.55 ± 0.32
4	6.69 ± 4.72	2.23 ± 1.31	1.40 ± 1.02	0.69 ± 0.49
5	5.22 ± 4.61	1.26 ± 0.91	1.50 ± 1.31	0.82 ± 0.71
6	7.42 ± 6.56	3.75 ± 2.61	...	1.72 ± 1.83
7	6.66 ± 6.46	2.73 ± 2.04	...	0.97 ± 0.70
8	9.82 ± 8.31	3.77 ± 2.46	...	1.70 ± 1.69
9	5.03 ± 3.79	2.14 ± 1.29	...	1.15 ± 0.78
10	5.42 ± 5.84	2.25 ± 1.80	...	1.06 ± 1.22

Ref. 13.) It can be observed that our method achieves the lowest mean registration errors. Table III shows the mean and STD on the 75 landmark points over all six phases by diffeomorphic Demons, $nD + t$ B-Splines, and our method. Again, our method achieves the lowest registration errors. It is worth noting that the maximum registration errors among 75 landmarks (from case #1 to case #5) are 6.08 mm by diffeomorphic Demons, 5.52 mm by $nD + t$ B-Splines, and 3.67 mm by our method, respectively.

The average intensity difference map between the reference image (MI phase image) and all registered phase images by diffeomorphic Demons is shown in Fig. 8(a). Similarly, the difference map between the aligned MI phase image and all other phase images in the common space by $nD + t$ B-Splines and our method are displayed in Figs. 8(b) and 8(c), respectively. According to the color map shown in the right of Fig. 8,

our method can align all phase images much better than the other two methods.

To evaluate the temporal consistency, we can further compute the velocity (i.e., the change ratio between two fiber landmarks from consecutive phases) at each respiratory phase. The average and standard deviation of velocity magnitude of 75 trajectories (corresponding to the landmarks used in Table III) in the first five cases of DIR-lab dataset are (0.85 ± 0.59) mm/phase by diffeomorphic Demons, (0.68 ± 0.33) mm/phase by $nD + t$ B-Splines method, and (0.72 ± 0.35) mm/phase by our method. Figures 9(a)–9(c) show the velocity magnitudes of a typical case in DIR-lab dataset, for all 75 trajectories, by diffeomorphic Demons, $nD + t$ B-Splines, and our method, respectively. The yellow curves denote the mean velocity magnitudes along respiratory phases. Since $nD + t$ B-Splines method and our method both constrain the continuity along respiratory phase, it can be observed that the velocity magnitudes by these two methods are much more continuous than diffeomorphic Demons.

Furthermore, for two of those 75 landmarks, we show their 3D motion trajectories in Fig. 10(b) and their superior–inferior projections by diffeomorphic Demons, $nD + t$ B-Splines, and our method in Fig. 10(a). The crosses in Fig. 10(c) designate the locations of the two selected landmark points in MI phase. It can be observed that the motion trajectories estimated by our method is not only much closer to the manual ground-truth trajectories, but also much smoother than those produced by other two methods.

III.B.2. Quality of estimated high-resolution lung images

Since our method is able to obtain high-resolution phase images by mapping the high-resolution group-mean image back onto each original phase image space, the discontinuous or missing vessels in the original phase images are

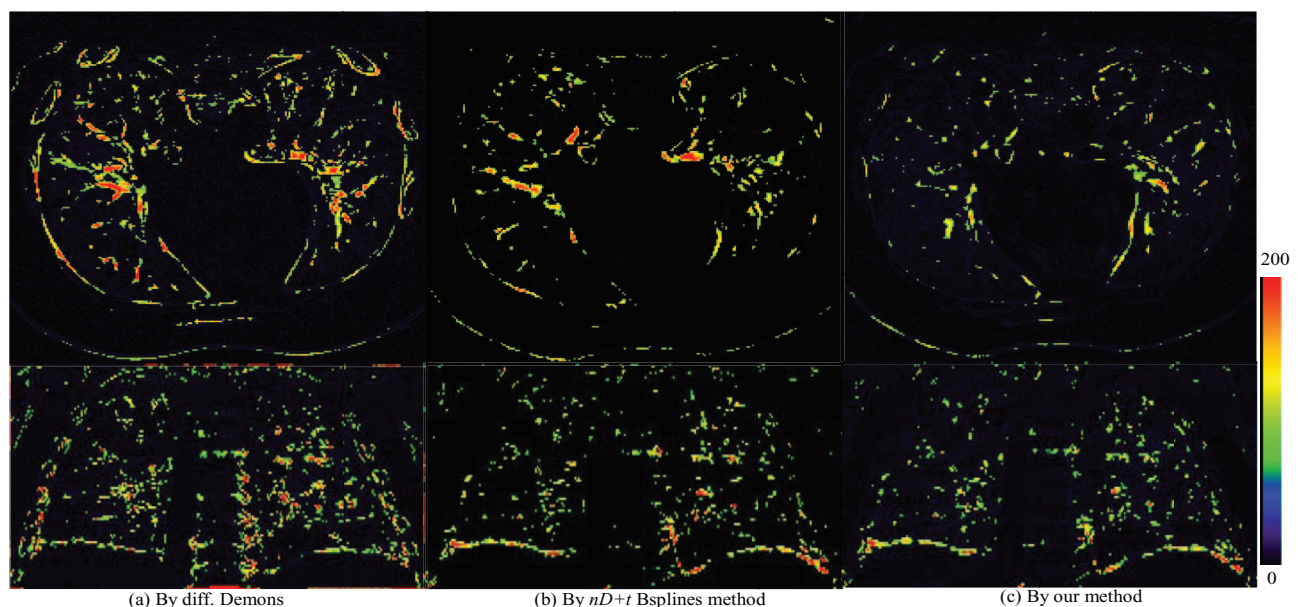


FIG. 8. The average intensity difference map after registration by diffeomorphic Demons, $nD + t$ B-Splines, and our method.

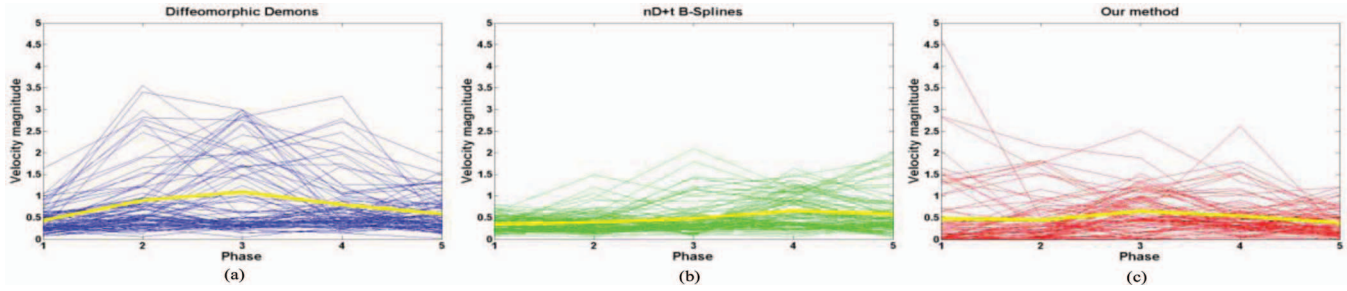


FIG. 9. The estimated velocity magnitudes from MI to ME phase for a typical 4D-CT.

expected to be revealed after performing our method. To demonstrate it, we calculate the Hessian map on each point to extract the lung vessels, as shown Fig. 11. The upper row in Fig. 11 shows the vessel trees from MI (phase #1) to ME (phase #6) in the original 4D-CT after linear interpolation along superior–inferior direction. Note that the vessels are ex-

tracted by setting threshold (top 10%) on the Hessian map and further using the morphological analysis method to discard the isolated points. By using the same procedure and same parameters, the vessel trees can also be extracted from the high-resolution phase images generated by our method as shown in the lower row of Fig. 11. It can be observed that

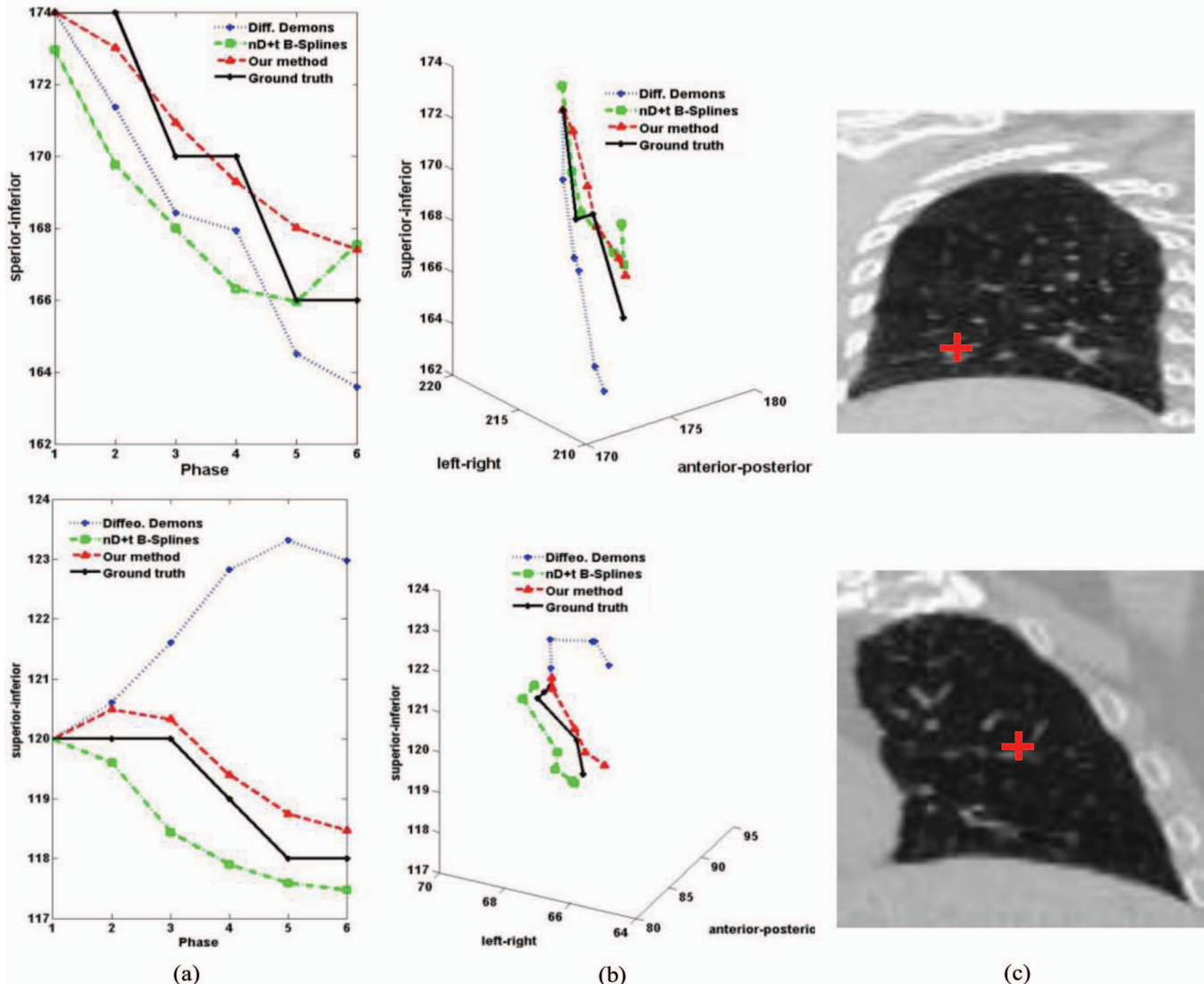


FIG. 10. The estimated 3D trajectories of two landmarks by Demons and our method.

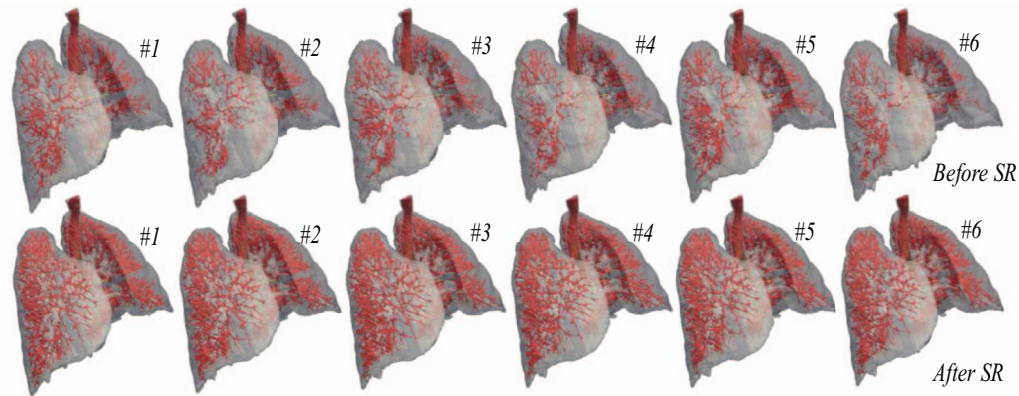


FIG. 11. The vessel trees in all respiratory phases before and after applying super-resolution.

(1) more anatomical details on vessels have been recovered by employing the super-resolution technique in our method, and (2) the vessel trees are temporally more consistent along the respiratory lung motion.

III.B.3. Computation time

All the experiments are performed on our DELL computation server with two CPUs (each with four 2.0 GHz cores) and 32G memory. For 4D-CTs (cases 1–5) from DIR-lab dataset with smaller image size (e.g., $256 \times 256 \times 99$), the computation time for diffeomorphic Demons, $nD + t$ B-Splines, and our proposed method is 0.5, 2.4, and 2.3 h, respectively. For the rest 4D-CTs (cases 6–10) with larger image size (e.g., $512 \times 512 \times 128$), the computation time by diffeomorphic Demons increases to 2 h, while $nD + t$ B-Splines and our proposed method have the similar computation time, with 2.5 and 2.4 h, respectively. Since diffeomorphic Demons equally treats each point, i.e., including background points, the computation time increases dramatically with the increased image size. On the contrary, $nD + t$ B-Splines method uses lung mask to confine the computation area, and our method hierarchically selects key points only from distinctive image regions. Thus, the computation loads by $nD + t$ B-Splines and our method do not increase dramatically as the image size increases.

III.C. Evaluation on lung cancer patient dataset

Additional lung 4D scans were acquired in UNC from 16-slice Philips Brilliance big bore helical CT scanner with 1.5 mm detector size per slice. Rotation period is 0.5 s and pitch is chosen so that couch does not move more than 1 detector length ($1.6 \text{ mm} \times 16 = 24 \text{ mm}$) in the time it takes to complete one full breath cycle. Respiratory signal is monitored through a bellow sensor and patient is instructed to breathe freely and normally. The phase images with respiratory cycle 0%, 10%, 20%, 30%, 40%, and 50% are reconstructed. The axial view of CT has 512×512 pixels (resolution is about 0.8 mm for a patient with 40 cm size) and the superior–inferior resolution is 3 mm. A sample of patient shown in Fig. 12(a) has a 20 mm tumor on the right lung.

III.C.1. Evaluation of respiratory motion

Given the estimated respiratory motion from 4D-CT, we are able to automatically propagate the manually labeled anatomical structures from one phase to all other phases, thus saving a lot of time in manually drawing contours in all phases. In this experiment, we first draw the contour of tumor in all phases, which are considered as ground truth. Then, we use diffeomorphic Demons, $nD + t$ B-Splines, and our method to map the contours in the maximum inhale phase to

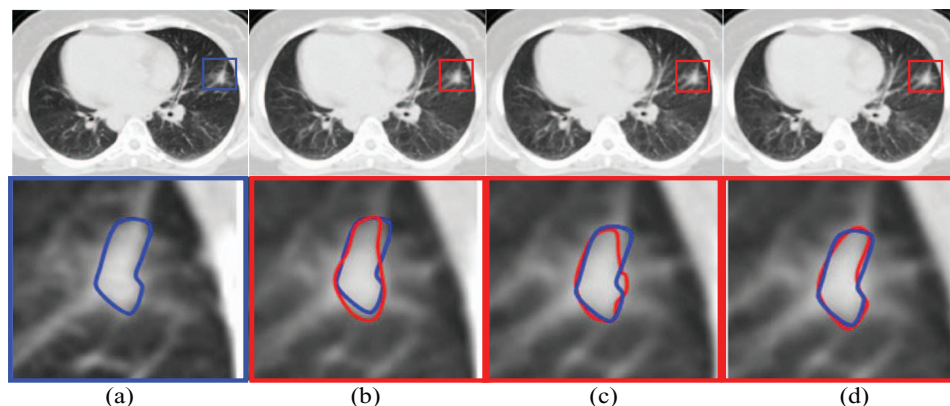


FIG. 12. The segmentation of tumor at maximum exhale phase.

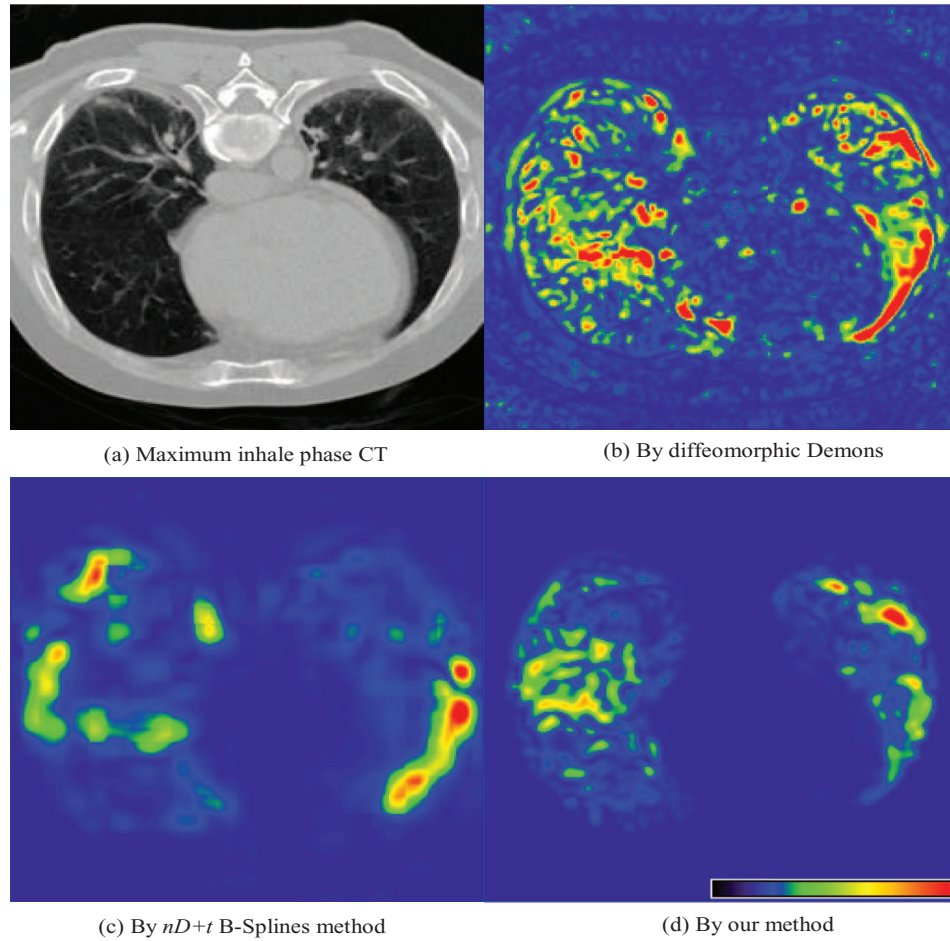


FIG. 13. The Laplacian map of resulting deformation fields by diffeomorphic Demons, $nD + t$ B-Splines and our method.

all other phases, by following the temporal deformation fields estimated by each method. Particularly, we set the phase image in maximum inhale stage as the reference image for diffeomorphic Demons. The propagated contours of tumor in the maximum exhale phase by diffeomorphic Demons, $nD + t$

B-Splines, and our method are shown in Figs. 12(b)–12(d), respectively, which are further overlaid with the manually labeled tumor [solid curve in Fig. 12(a)]. Compared with the ground-truth tumor, our method achieves the best labeling result against other two methods.

TABLE IV. The mean and standard deviation of registration errors (mm) on 300 landmark points between MI and ME phases in evaluating each component of our motion estimation method. Bold is used for emphasis that the residual errors by our method are significantly smaller than other counterparts.

#	300 landmark points between MI and ME			
	Our method without groupwise registration	Our method without temporal smoothing	Our method without super-resolution	Our full method
1	0.86 ± 0.95	0.69 ± 0.78	0.85 ± 0.82	0.64 ± 0.61
2	0.90 ± 1.06	0.63 ± 0.80	0.79 ± 0.65	0.56 ± 0.63
3	1.05 ± 1.65	0.72 ± 0.75	0.92 ± 0.54	0.70 ± 0.68
4	1.35 ± 1.71	1.01 ± 1.26	1.12 ± 0.82	0.91 ± 0.79
5	1.69 ± 1.23	1.26 ± 0.99	1.43 ± 0.96	1.10 ± 1.14
6	5.02 ± 2.06	3.93 ± 3.50	6.95 ± 4.06	3.28 ± 3.45
7	2.21 ± 2.07	1.75 ± 2.72	3.64 ± 2.15	1.68 ± 1.22
8	2.27 ± 2.13	2.07 ± 2.13	4.05 ± 2.64	1.70 ± 1.69
9	1.91 ± 1.57	2.11 ± 1.43	3.96 ± 1.85	1.72 ± 1.32
10	1.83 ± 1.92	1.93 ± 2.00	3.25 ± 2.68	1.48 ± 1.84

TABLE V. The mean and standard deviation of registration errors (mm) on 75 landmark points from MI to ME phases in evaluating each component of our motion estimation method. Bold is used for emphasis that the residual errors by our method are significantly smaller than other counterparts.

#	75 landmark points across all six phases			
	Our method without groupwise registration	Our method without temporal smoothing	Our method without super-resolution	Our full method
1	0.91 ± 0.64	0.66 ± 0.53	0.79 ± 0.46	0.51 ± 0.39
2	1.06 ± 1.11	0.61 ± 0.46	0.86 ± 0.52	0.47 ± 0.34
3	0.85 ± 0.56	0.77 ± 0.47	0.92 ± 0.53	0.55 ± 0.32
4	1.30 ± 1.13	0.89 ± 0.61	1.07 ± 0.75	0.69 ± 0.49
5	0.98 ± 0.85	0.98 ± 0.88	1.13 ± 0.86	0.82 ± 0.71
6	2.31 ± 1.90	2.00 ± 2.49	2.26 ± 1.96	1.72 ± 1.83
7	1.96 ± 2.03	1.46 ± 1.03	1.85 ± 0.88	0.97 ± 0.70
8	2.43 ± 2.11	1.95 ± 1.96	2.40 ± 1.95	1.70 ± 1.69
9	1.93 ± 1.19	1.58 ± 1.51	2.05 ± 1.15	1.15 ± 0.78
10	1.81 ± 1.18	1.33 ± 1.37	1.69 ± 1.52	1.06 ± 1.22

III.C.2. Evaluation on the deformation fields

We further evaluate the smoothness of the estimated deformation fields between two extreme phases by diffeomorphic Demons, $nD + t$ B-Splines, and our method. Specifically, we calculate the Laplacian value of displacement vector at each point. Larger degree of Laplacian value indicates less smoothness in the deformation field. The average Laplacian degrees (excluding background points) are 1.04 by diffeomorphic Demons, 0.49 by $nD + t$ B-Splines method, and 0.26 by our method. Figures 13(b)–13(d) show the Laplacian maps by diffeomorphic Demons, $nD + t$ B-Splines, and our method, respectively, where greater brightness indicates larger Laplacian value. It can be observed that the resulted deformation field by $nD + t$ B-Splines and our method achieve similar smooth deformation fields, while the deformation field by diffeomorphic Demons is less favored.

III.D. Discussion

As mentioned above, our method overcomes the limitations of the conventional motion estimation methods in lung 4D-CT in three ways: (1) we use the concept of groupwise registration to avoid the bias in selecting the reference image during registration; (2) we perform temporal smoothing along temporal fibers to achieve temporal consistency; and (3) we construct the super-resolution group-mean image to improve the accuracy of correspondence detection. In order to comprehensively evaluate the contribution of each component, we compare our full method with respect to three degraded methods: (1) our method without using groupwise registration by deforming all phase images to the reference image (i.e., maximum inhale phase image); (2) our method without temporal smoothing by skipping the temporal smoothing step; and (3) our method without using super-resolution group-mean image. Similar to Sec. III.B, we apply these three degraded methods to the ten cases of 4D-CT in DIR-lab dataset. The mean and standard deviation of registration errors on the 300

landmarks between MI and ME phases and the 75 landmarks from MI to ME phases are provided in Tables IV and V, respectively. It is obvious that the accuracy of motion estimation decreases significantly after removing one component from our full method.

IV. CONCLUSION

In this paper, a novel motion estimation method has been presented to measure lung respiratory motion in 4D-CT. Our method is implemented by repeating the two iterative steps, i.e., (1) simultaneously aligning all phase images onto the common space by spatiotemporal registration, and (2) reconstructing the high-resolution group-mean image with improved anatomical details by super-resolution technique. We have comprehensively evaluated our proposed method on both simulated dataset and real patient 4D-CT dataset, where the 4D-CTs have large interslice thickness. Compared to other state-of-the-art registration methods, our algorithm improves not only the accuracy of motion estimation, but also the temporal consistency of estimated lung motion.

ACKNOWLEDGMENT

The authors thank Dr. Metz, the corresponding author of Ref. 13, to kindly provide us the estimated motion trajectories.

APPENDIX: SOLUTION TO WEIGHTING VECTOR

We assume all $D_t \in \mathcal{D}$ and $\tilde{R}_G(x)$ are arranged into vectors. Given $\mathbf{\Lambda}^K = \{D_{x(q)} | q = 1, \dots, K\}$ with K elements, the optimization of weighting vector $\mathbf{w}(\{\mathbf{\Lambda}^K\})$ in Eq. (14) consists of three steps:⁵⁶

- (1) Calculate the $K \times K$ correlation matrix A , with A_{ij} defined as the inner production of $D_{x(i)}$ and $D_{x(j)}$. A^{-1} is the inverse of matrix A .
- (2) Compute the Lagrange multiplier $\lambda = alb$, where $a = 1 - \sum_{ij} A_{ij}^{-1}(\tilde{R}_G(x) \cdot D_{x(j)})$ and $b = \sum_{ij} A_{ij}^{-1}$.

(3) The weight $w_{x(j)}^k$ for reconstruction is computed as

$$w_i = \sum_j A_{ij}^{-1} (R_G(x) \cdot D_{x(j)} + \lambda).$$

^{a)}Electronic mail: grwu@med.unc.edu

^{b)}Electronic mail: qianwang@cs.unc.edu

^{c)}Author to whom correspondence should be addressed. Electronic mail: jun_lian@med.unc.edu

^{d)}Electronic mail: dgshen@med.unc.edu

¹E. Rietzel, T. Pan, and G. T. Y. Chen, "Four-dimensional computed tomography: Image formation and clinical protocol," *Med. Phys.* **32**, 874–889 (2005).

²P. Keall, G. Mageras, J. Balter, R. Emery, K. Forster, S. Jiang, J. Kapatoes, D. Low, M. Murphy, B. Murray, C. Ramsey, M. Van Herk, S. Vedam, J. Wong, and E. Yorke, "The management of respiratory motion in radiation oncology report of AAPM Task Group 76," *Med. Phys.* **33**, 3874–3900 (2006).

³T. Zhang, N. P. Orton, and W. A. Tomé, "On the automated definition of mobile target volumes from 4D-CT images for stereotactic body radiotherapy," *Med. Phys.* **32**, 3493–3503 (2005).

⁴X. Zhang, K.-L. Zhao, T. M. Guerrero, S. E. McGuire, B. Yaremko, R. Komaki, J. D. Cox, Z. Hui, Y. Li, W. D. Newhauser, R. Mohan, and Z. Liao, "4D CT-based treatment planning for intensity-modulated radiation therapy and proton therapy for distal esophagus cancer," *Int. J. Radiat. Oncol., Biol., Phys.* **72**, 278–287 (2006).

⁵F. Khan, G. Bell, J. Antony, M. Palmer, P. Balter, K. Bucci, and M. J. Chapman, "The use of 4D CT to reduce lung dose: A dosimetric analysis," *Med. Dosim.* **34**, 273–278 (2008).

⁶A. Trofimov, E. Rietzel, H.-M. Lu, B. Martin, S. Jiang, G. T. Y. Chen, and T. Bortfeld, "Temporo-spatial IMRT optimization: Concepts, implementation and initial results," *Phys. Med. Biol.* **50**, 2779–2798 (2005).

⁷T. Guerrero, G. Zhang, W. Segars, T.-C. Huang, S. Bilton, G. Ibbott, L. Dong, K. Forster, and K. P. Lin, "Elastic image mapping for 4-D dose estimation in thoracic radiotherapy," *Radiat. Protect. Dosim.* **115**, 497–502 (2005).

⁸T. Guerrero, K. Sanders, E. Castillo, Y. Zhang, L. Bidaut, T. Pan, and R. Komaki, "Dynamic ventilation imaging from four-dimensional computed tomography," *Phys. Med. Biol.* **51**, 777–791 (2006).

⁹S. Rit, D. Sarrut, and L. Desbat, "Comparison of analytic and algebraic methods for motion-compensated cone-beam CT reconstruction of the thorax," *IEEE Trans. Image Process.* **28**, 1513–1526 (2009).

¹⁰J. R. McClelland, J. M. Blackall, S. Tarte, and D. J. Hawkes, "A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy," *Med. Phys.* **33**, 3348–3359 (2006).

¹¹R. Castillo, E. Castillo, R. Guerra, V. E. Johnson, T. McPhail, A. K. Garg, and T. Guerrero, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Phys. Med. Biol.* **54**, 1849–1870 (2009).

¹²J. Ehrhardt, R. Werner, A. Schmidt-Richberg, and H. Handels, "Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration," *IEEE Trans. Med. Imaging* **30**, 251–265 (2011).

¹³C. T. Metz, S. Klein, M. Schaap, T. van Walsum, and W. J. Niessen, "Non-rigid registration of dynamic medical imaging data using nD+t B-splines and a groupwise optimization approach," *Med. Image Anal.* **15**, 238–249 (2011).

¹⁴E. Castillo, R. Castillo, J. Martinez, M. Shenoy, and T. Guerrero, "Four-dimensional deformable image registration using trajectory modeling," *Phys. Med. Biol.* **55**, 305–327 (2010).

¹⁵G. E. Christensen, J. H. Song, W. Lu, I. E. Naqa, and D. A. Low, "Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry," *Med. Phys.* **34**(6), 2155–2163 (2007).

¹⁶J. Ehrhardt, R. Werner, D. Säring, T. Frenzel, W. Lu, D. Low, and H. Handels, "An optical flow based method for improved reconstruction of 4D CT data sets acquired during free breathing," *Med. Phys.* **34**, 711–721 (2007).

¹⁷Y. Yin, E. Hoffman, and C. Lin, "Mass preserving nonrigid registration of CT lung images using cubic B-spline," *Med. Phys.* **36**, 4213–4222 (2009).

¹⁸J. Vandemeulebroucke, S. Rit, D. Sarrut, J. Kybic, and P. Clarysse, "Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs," *Med. Phys.* **38**, 166–179 (2011).

¹⁹N. J. Tustison, S. P. Awate, S. Gang, T. S. Cook, and J. C. Gee, "Point set registration using Havrda-Charvat-Tsallis entropy measures," *IEEE Trans. Med. Imaging* **30**, 451–460 (2010).

²⁰R. Colgan, J. McClelland, D. McQuaid, P. M. Evans, D. Hawkes, J. Brock, D. Landau, and S. Webb, "Planning lung radiotherapy using 4D CT data and a motion model," *Phys. Med. Biol.* **53**, 5815–5830 (2008).

²¹V. Gorbunova, S. Durrleman, L. Pechin, X. Pennec, and M. de Bruijne, "Lung CT registration combining intensity, curves and surfaces," in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (IEEE Conference Publications, Congress Center "De Doelen", Rotterdam, The Netherlands, 2010), pp. 340–343.

²²X. Han, "Feature-constrained nonlinear registration of lung CT images," in *EMPIRE10: MICCAI 2010 Workshop on Evaluation of Methods for Pulmonary Image Registration*, Beijing, 2010.

²³V. Boldea, G. Sharp, S. Jiang, and D. Sarrut, "4D-CT lung motion estimation with deformable registration: Quantification of motion nonlinearity and hysteresis," *Med. Phys.* **35**, 1008–1018 (2008).

²⁴G. Székely, H. Hahn, M. Yigitsoy, C. Wachinger, and N. Navab, "Temporal groupwise registration for motion modeling," in *Information Processing in Medical Imaging* (Springer, Kloster Irsee, Germany, 2011), pp. 648–659.

²⁵R. Li, J. H. Lewis, L. I. Cerviño, and S. B. Jiang, "4D CT sorting based on patient internal anatomy," *Phys. Med. Biol.* **54**, 4821–4833 (2009).

²⁶M. Georg, R. Souvenir, A. Hope, and R. Pless, "Manifold learning for 4D CT reconstruction of the lung," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, CVPRW'08* (IEEE Conference Publications, Anchorage, Alaska, 2008), pp. 1–8.

²⁷T. Yamamoto, U. Langner, B. W. Loo Jr, J. Shen, and P. J. Keall, "Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients," *Int. J. Radiat. Oncol., Biol., Phys.* **72**, 1250–1258 (2008).

²⁸J. H. Lewis and S. B. Jiang, "A theoretical model for respiratory motion artifacts in free breathing CT scans," *Phys. Med. Biol.* **54**, 745–755 (2009).

²⁹E. Johnston, M. Diehn, J. D. Murphy, J. Billy, W. Loo, and P. G. Maxim, "Reducing 4D CT artifacts using optimized sorting based on anatomic similarity," *Med. Phys.* **38**, 2424–2429 (2011).

³⁰R. George, T. D. Chung, S. S. Vedam, V. Ramakrishnan, R. Mohan, E. Weiss, and P. J. Keall, "Audio-visual biofeedback for respiratory-gated radiotherapy: Impact of audio instruction and audio-visual biofeedback on respiratory-gated radiotherapy," *Int. J. Radiat. Oncol., Biol., Phys.* **65**, 924–933 (2006).

³¹R. B. Venkat, A. Sawant, Y. Suh, R. George, and P. J. Keall, "Development and preliminary evaluation of a prototype audiovisual biofeedback device incorporating a patient-specific guiding waveform," *Phys. Med. Biol.* **53**, 197–208 (2008).

³²T. Pan, X. Sun, and D. Luo, "Improvement of the cine-CT based 4D-CT imaging," *Med. Phys.* **34**, 4449–4454 (2007).

³³A. F. Abdelnour, S. A. Nehmeh, T. Pan, J. L. Humm, P. Vernon, H. Schöder, K. E. Rosenzweig, G. S. Mageras, E. Yorke, S. M. Larson, and Y. E. Erdi, "Phase and amplitude binning for 4D-CT imaging," *Phys. Med. Biol.* **52**, 3515–3529 (2007).

³⁴R. Zeng, J. A. Fessler, J. M. Balter, and P. A. Balter, "Iterative sorting for four-dimensional CT images based on internal anatomy motion," *Med. Phys.* **35**, 917–927 (2008).

³⁵D. Han, J. Bayouth, S. Bhatia, M. Sonka, and X. Wu, "Characterization and identification of spatial artifacts during 4D-CT imaging," *Med. Phys.* **38**, 2074 (2011).

³⁶E. Schreibmann, G. T. Y. Chen, and L. Xing, "Image interpolation in 4D CT using a BSpline deformable registration model," *Int. J. Radiat. Oncol., Biol., Phys.* **64**, 1537–1550 (2006).

³⁷R. Werner, J. Ehrhardt, T. Frenzel, D. Säring, W. Lu, D. Low, and H. Handels, "Motion artifact reducing reconstruction of 4D CT image data for the analysis of respiratory dynamics," *Methods Inf. Med.* **46**, 254–260 (2007).

³⁸G. Wu, Q. Wang, and D. Shen, "Registration of longitudinal brain image sequences with implicit template and spatial-temporal heuristics," *NeuroImage* **59**, 404–421, 2012.

³⁹S. Mori and P. C. M. v. Zijl, "Fiber tracking: Principles and strategies—A technical review," *NMR Biomed.* **15**, 468–480 (2002).

⁴⁰M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction," *IEEE Trans. Image Process.* **18**, 36–51 (2009).

⁴¹J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.* **19**, 2861–2873 (2010).

- ⁴²S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Trans. Image Process.* **13**, 1327–1344 (2004).
- ⁴³T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage* **45**, S61–S72 (2009).
- ⁴⁴J. Hamm, C. Davatzikos, and R. Verma, "Efficient large deformation registration via geodesics on a learned manifold of images," *Med. Image Comput. Comput. Assist. Interv.* **12**, 680–687 (2009).
- ⁴⁵H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Underst.* **89**, 114–141 (2003).
- ⁴⁶S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Trans. Med. Imaging* **20**, 490–498 (2001).
- ⁴⁷A. Frangi, W. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98* (Springer, Berlin/Heidelberg, 1998), Vol. 1496, pp. 130–137.
- ⁴⁸C. H. Lo and H. S. Don, "3-D moment forms: Their construction and application to object identification and positioning," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 1053–1064 (1989).
- ⁴⁹G. Wu, P.-T. Yap, M. Kim, and D. Shen, "TPS-HAMMER: Improving HAMMER registration algorithm by soft correspondence matching and thin-plate splines based deformation interpolation," *NeuroImage* **49**, 2225–2233 (2010).
- ⁵⁰D. Shen, "Fast image registration by hierarchical soft correspondence detection," *Pattern Recognit.* **42**, 954–961 (2009).
- ⁵¹F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 567–585 (1989).
- ⁵²M. P. Wand and M. C. Jones, *Kernel Smoothing, Series Monographs on Statistics and Applied Probability* (Chapman and Hall, New York, 1995).
- ⁵³E. A. Nadaraya, "On estimating regression," *Theory Probab. Appl.* **9**, 157–159 (1964).
- ⁵⁴Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Trans. Image Process.* **19**, 1153–1165 (2011).
- ⁵⁵G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approx.* **13**, 57–98 (1997).
- ⁵⁶S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science* **290**, 2323–2326 (2000).
- ⁵⁷S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging* **29**, 196–205 (2010).
- ⁵⁸D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imaging* **18**, 712–721 (1999).
- ⁵⁹In the scenario of groupwise registration, each phase CT is considered to be sitting in a high-dimensional image manifold, where each phase CT image is considered as a point in a space. Group-mean image is defined as a point in the manifold that has the lowest overall distance to all phase CTs. The space where the group-mean image sits is also called as common space in the literature for groupwise registration.